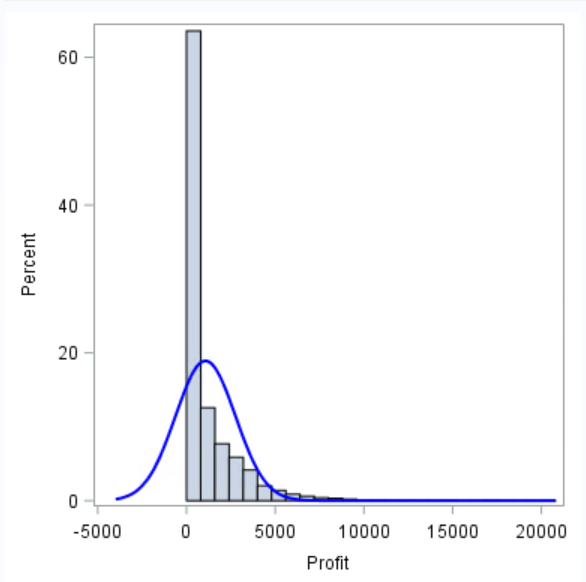
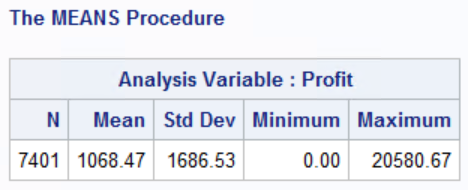
**HW6- Group -10- Shweta Siddha, Manisha Gupta, Achint Khanijo, Kartikay Nigam**

1. **Run the following Tobit model (Use PROC QLIM)**

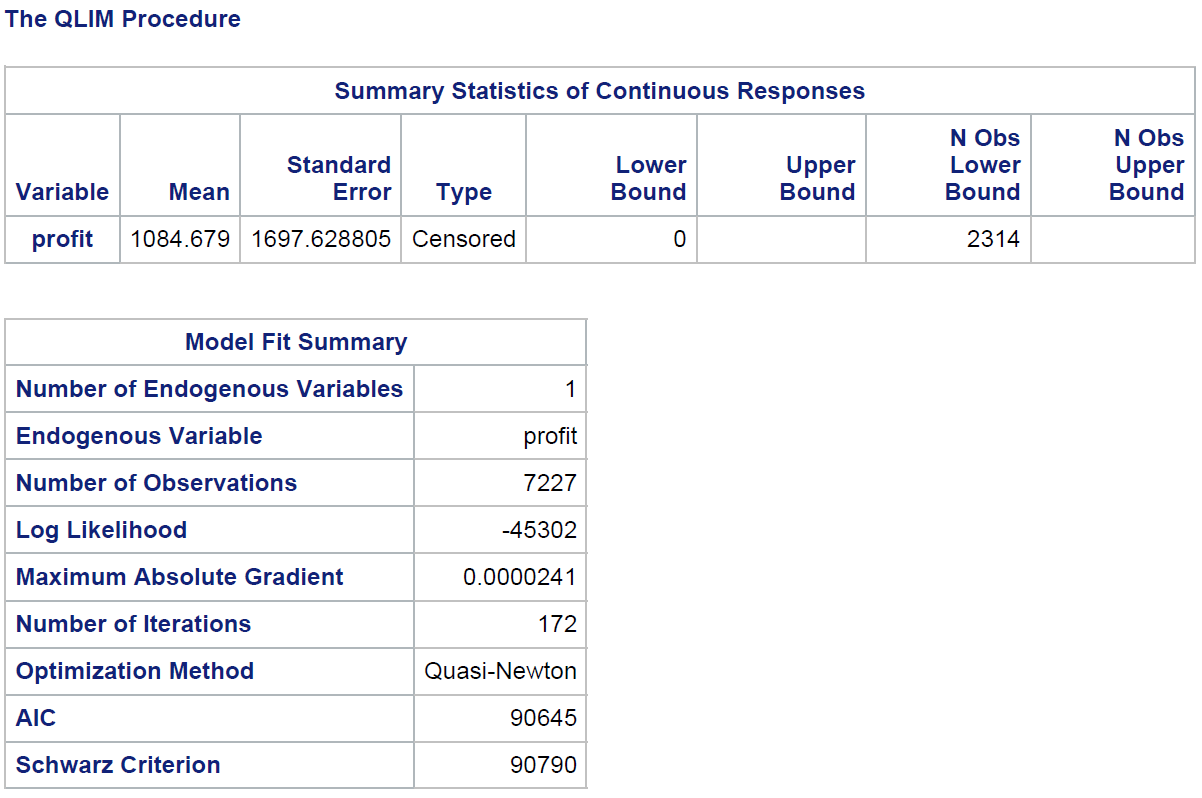
**Model profit = age, totaltrans, rewards, limit, numcard, modes of acquisition, type of card, types of affinity**

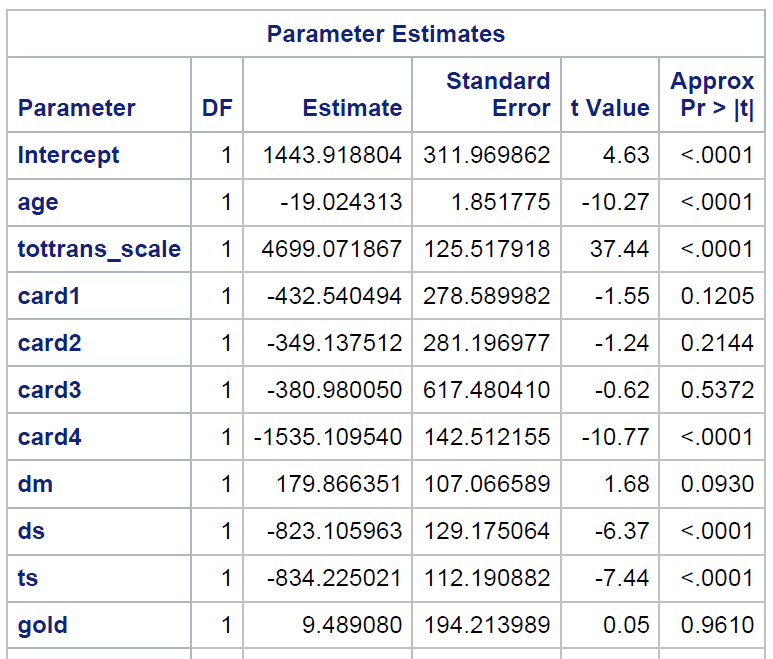
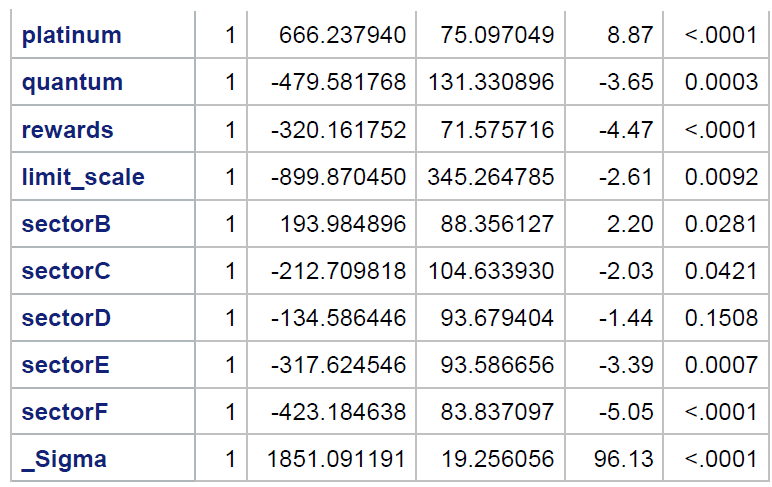
**Write a summary of the results. Focus on important effects, interpretation, model fit etc**

**Analysis of Profit column**

Also looking at the above histogram showing the distribution of Profit column, we can see the censoring in the data, there are many values with Profit = 0.



**Model Fit Summary:**

Total number of observations: 7227

Log likelihood: -45302

**Age:** One-year increase in **age** of the customer is associated with approx. 19 units decrease in profit over a 3-year period, keeping all other variables constant. And this impact is statistically significant.

**Total transaction**: 100000 units increase in **Total transaction** is associated with approx. 4699 units increase in profit over a 3-year period, keeping all other variables constant. And this impact is statistically significant.

**Number of Card 0:** Reference category

**Number of Card 1:** Customer who has one card is not statistically significant as compared to person having no card.

**Number of Card 2:** Customer who has two card is not statically significant as compared to person having no card.

**Number of Card 3 -** Customer who has three card is not statically significant as compared to person having no card.

**Number of Card 4:** Profit is 1535.1 units less from customers with four cards from this bank as compared to those who have 0, 1, 2 or 3cards.

**NET**: Customer acquired through net is taken as reference category.

**DM:** Profit is 179.86 units more for customers who are acquired via direct mail as compared to those who are acquired through internet, keeping all other variables constant. It is not significant at 95% confidence interval as its p value is above the threshold value of 5%. It can be significant at around 90% confidence level.

**TS:** Profit is 834.22 units less for customers who are acquired via telephone selling as compared to those who are acquired through internet, keeping all other variables constant. And this impact is statistically significant.

**DS:** Profit is 823.10 units less for customers who are acquired via direct selling as compared to those who are acquired through internet, keeping all other variables constant. And this impact is statistically significant.

**Standard:** whether the customer has standard card is taken as Reference category

**Gold**: There is no significant difference in Total finance charges paid by a customer over a 3-year period between customers with Standard card & Gold card.

**Platinum**: Profit is 666.237 units more for customers who have PLATINUM card as compared to customers who have STANDARD card, keeping all other variables constant. And this impact is statistically significant

**Quantum:** Profit is 479.58 dollar less for customers who have QUANTUM card as compared to customers who have STANDARD card, keeping all other variables constant. And this impact is statistically significant.

**Reward card**: Customer having **reward card** are associated with 320.16 dollar less profit as compared to customer who do not have reward cards, keeping all other variables constant. And this impact is statistically significant.

**Limit:** Every 100000 dollar increase in limit of a card is associated with 899.87 dollar decrease in profit, keeping all other variables constant. And this impact is statistically significant.

**SectorA:** No affinity card is taken as reference category.

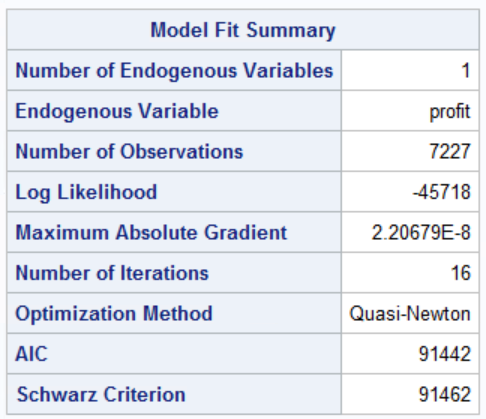
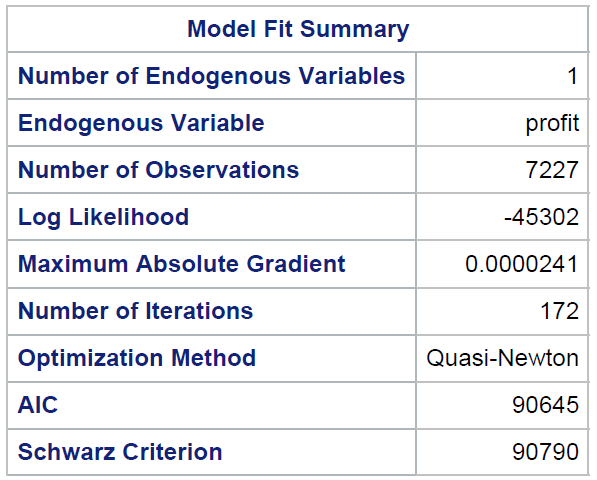
**SectorB:** profit is 193 dollar more for Affinity card affiliated with Professional organization as compared to customers who have No affinity card, keeping all other variables constant. And this impact is statistically significant

**SectorC:** profit is 212.71 dollar less for Affinity card affiliated with Sports as compared to customers who have No affinity card, keeping all other variables constant. And this impact is statistically significant.

**SectorD:** profit is 134.58 dollar less for Affinity card affiliated with Financial institution as compared to customers who have No affinity card, keeping all other variables constant. And this impact is not statistically significant

**SectorE:** profit is 317.62 dollar less for Affinity card affiliated with University as compared to customers who have No affinity card, keeping all other variables constant. And this impact is statistically significant.

**SectorF:** Profit is 423.18 dollar less for Affinity card affiliated with Commercial as compared to customers who have No affinity card. And this impact is statistically significant.

**Model Fit Statistics:**

**AIC**: This is the Akaike Information Criterion. AIC, like Adjusted R-square in linear regression, penalize the log-likelihood for the number of predictors in the model. Ultimately, the model with the smallest AIC is considered good.

**SC**: This is the Schwarz Criterion. Like AIC, SC penalizes for the number of predictors in the model and the smallest SC is most desirable.

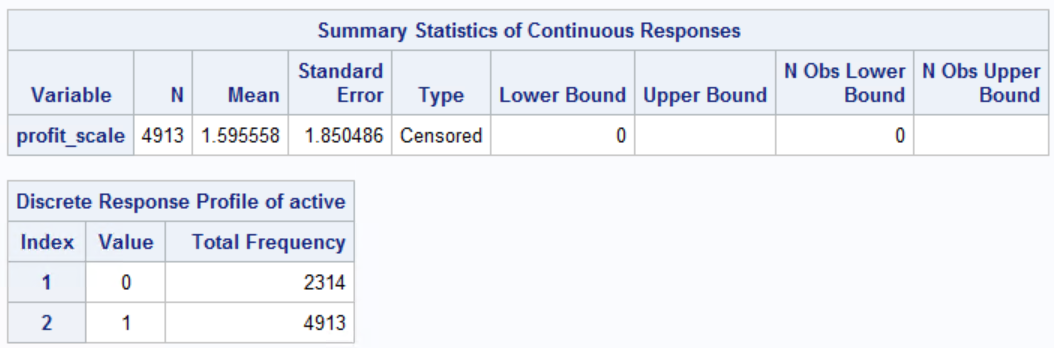
As we cannot compare the AIC and SC of model with only intercept against model with covariates, we are testing it for model with one parameter against model with all important variables. We can observe that AIC and SC for the model with only one explanatory variable limit is 91442 and 91462 which significantly reduces to 90645 and 90790 respectively after adding all the relevant explanatory variables, so we can say that our model does better.

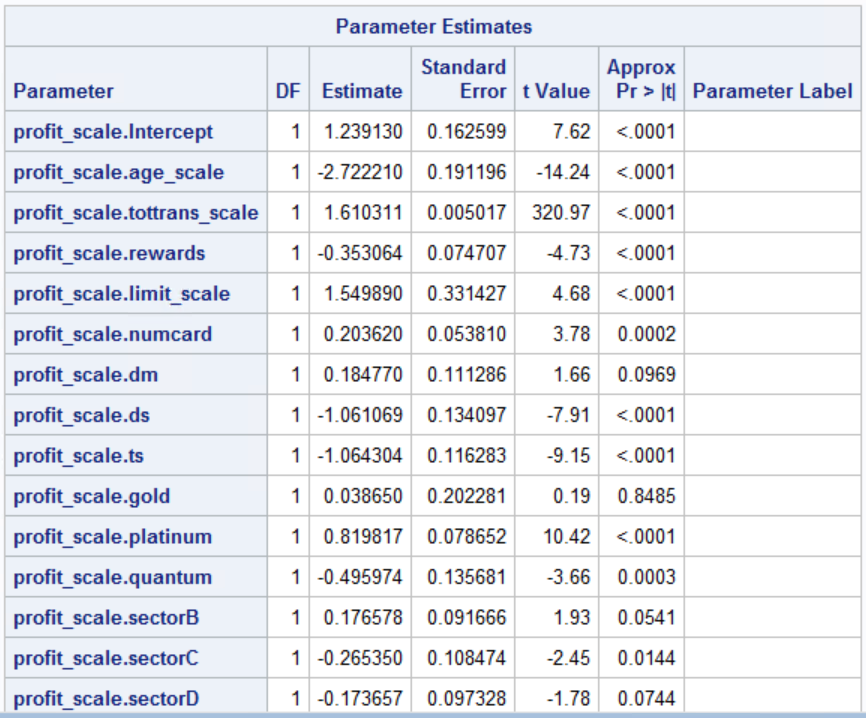
**Q2. Run a selection model (Use PROC QLIM)**

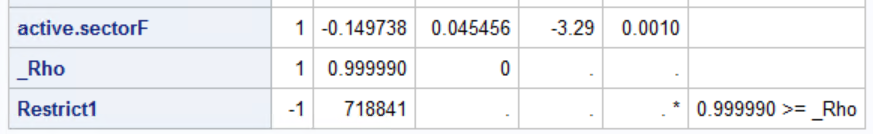
**Model active = age, rewards, limit, numcard, modes of acquition, type of card, types of affinity**

**Model profit = age, totaltrans, rewards, limit, numcard, modes of acquition, type of card, types of affinity**

**Write a summary of the results. Focus on important effects, interpretation, model fit etc**.





**Variable Interpretation:**

As per the selection model results, 4913 out of 7401 customers who have credit cards are active users. If the selection is not considered then the total profit tend to give biased values. The selection model avoids the selection bias by taking only the active users out of all the credit-card holders.

**Variable Interpretation wr.r.t active (at 5% significance level: p-value<0.05):**

The significant variables are total transaction amount, credit limit, rewards, numcard, ds, ts, platinum, quantum, sectorC, sectorB, sectorE, sectorF, age.

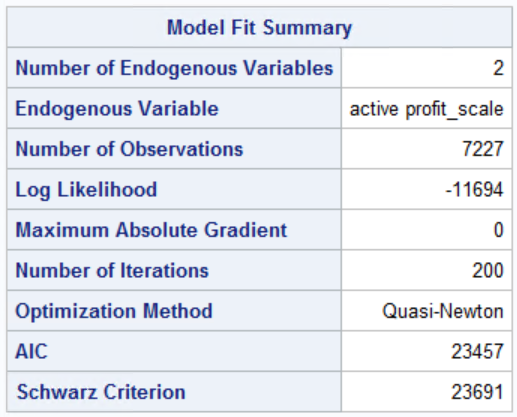
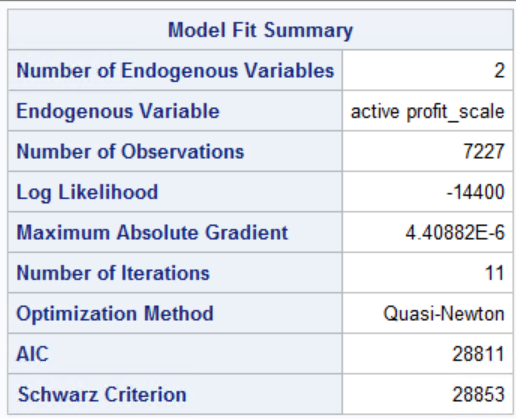
* One-year increase in age of the customer is associated with approx. $27.6 decrease in profit over a 3-year period. And this impact is statistically significant.
* $100000 increase in Total transaction amount spent by a customer in 3-year period is associated with approx. $39 increase in Profit. And this impact is statistically significant.
* Profit over a 3-year period are approx. $403 less if the customer has a Rewards card as compared to the customer who doesn’t have a Rewards card. And this impact is statistically significant.
* $100000 increase in Credit limit of customer is associated with approx. $226 increase in profit over a 3-year period. And this impact is statistically significant.
* 1 unit increase in number of cards owned by a customer is associated with $ 179 increase in profit over a 3-year period. And this impact is statistically significant.
* Profit over a 3-year period are $295 more if the customer’s mode of acquisition is ‘Direct Mail’ as compared to the customer acquired through ‘Internet’. And this impact is statistically significant
* Profit over a 3-year period are $916 less if the customer’s mode of acquisition is ‘Direct Selling’ as compared to the customer acquired through ‘Internet’. And this impact is statistically significant
* Profit over a 3-year period are $942 less if the customer’s mode of acquisition is ‘Telephone Selling’ as compared to the customer acquired through ‘Internet’. And this impact is statistically significant
* There is no significant difference in profit over a 3-year period between customers with Standard card & Gold card.
* Profit over a 3-year period are $726 more if the customer has a ‘Platinum card’ as compared to the customer who has a ‘Standard card’.
* Profit over a 3-year period are $793 less if the customer has a ‘Quantum card’ as compared to the customer who has a ‘Standard card’.
* Profit over a 3-year period less for the ones with Affinity cards as compared to the customers who have non-affinity cards.
* Profit over a 3-year period is $176 less for the ones with affiliated with Professional organization as compared to the customers who have non-affinity cards.
* Profit over a 3-year period is $265 less for the ones with affiliated with Sports as compared to the customers who have non-affinity cards.
* Profit over a 3-year period is $174 less for the ones with affiliated with Financial institution as compared to the customers who have non-affinity cards. however, this is not statically significant with 95% confidence level.
* Profit over a 3-year period is $409 less for the ones with affiliated with University as compared to the customers who have non-affinity cards.
* Profit over a 3-year period is $294 less for the ones with affiliated with Commercial as compared to the customers who have non-affinity cards.
* The probability of a customer being active decreases with increase in customer’s age.
* The probability of a customer being active is less for Rewards card customer as compared to non-reward card customer.
* The probability of a customer being active increases with increase in customer’s credit limit.
* The probability of a customer being active increases with increase in number of cards customer has from the firm.
* The probability of a customer being active is less for customers acquired through Direct Selling & Telephone selling as compared to customers acquired through Internet.
* The probability of a customer being active is more for customers having Platinum card as compared to customers having standard card.
* The probability of a customer being active is less for customers having Quantum card as compared to customers having standard card.
* The probability of customer being active is less for Affinity card (Sector C & Sector E) holders as compared to non-affinity card holders.

**Comparison of AIC and SC**

**AIC**: This is the Akaike Information Criterion. AIC, like Adjusted R-square in linear regression, penalize the log-likelihood for the number of predictors in the model. Ultimately, the model with the smallest AIC is considered good.

**SC**: This is the Schwarz Criterion. Like AIC, SC penalizes for the number of predictors in the model and the smallest SC is most desirable.

As we cannot compare the AIC and SC of model with only intercept against model with covariates, we are testing it for model with one parameter against model with all important variables. We can observe that AIC and SC for the model with only one explanatory variable limit is 28811 and 28853 which significantly reduces to 23457 and 23691 respectively after adding all the relevant explanatory variables, so we can say that our model does better.

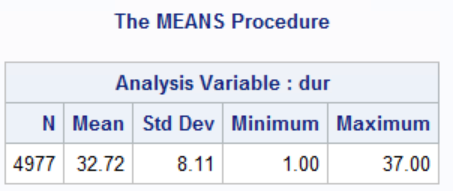
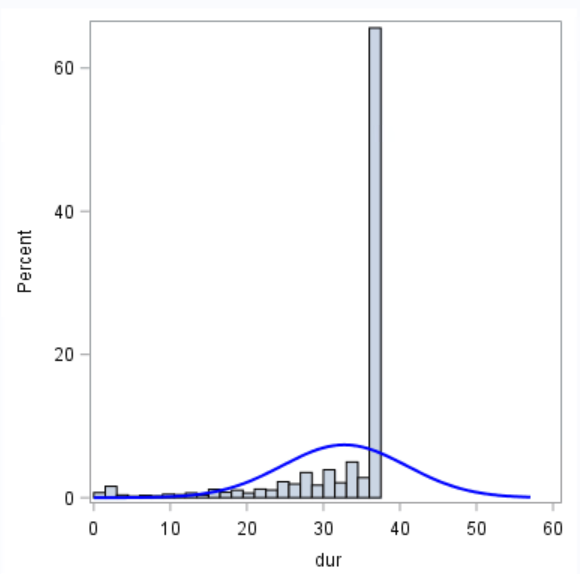
 

**Q3: Survival analysis**

**1. Delete all customers who are inactive.**

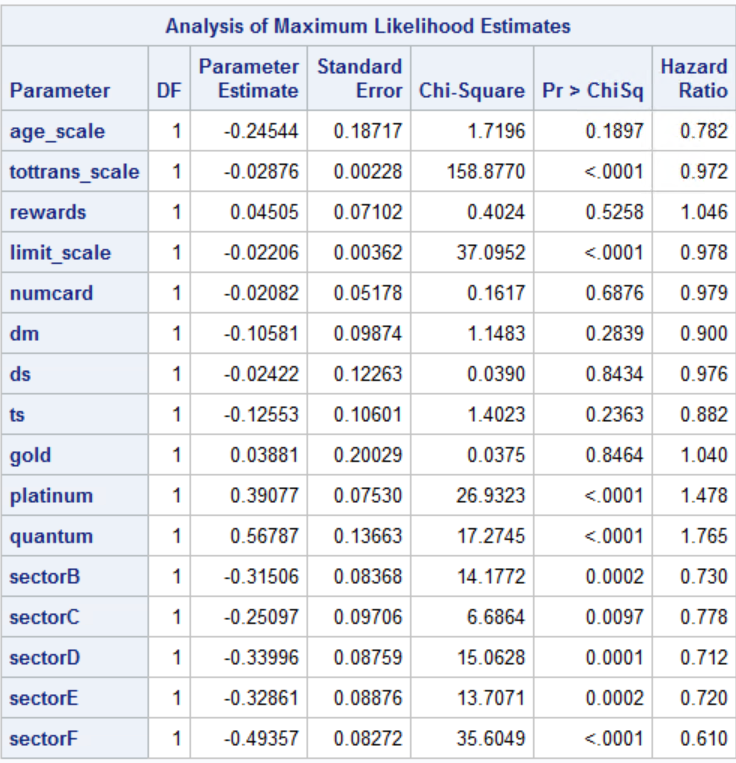
**2. Run a proportional hazards model (PROC PHREG)**

**Analysis of duration column:**

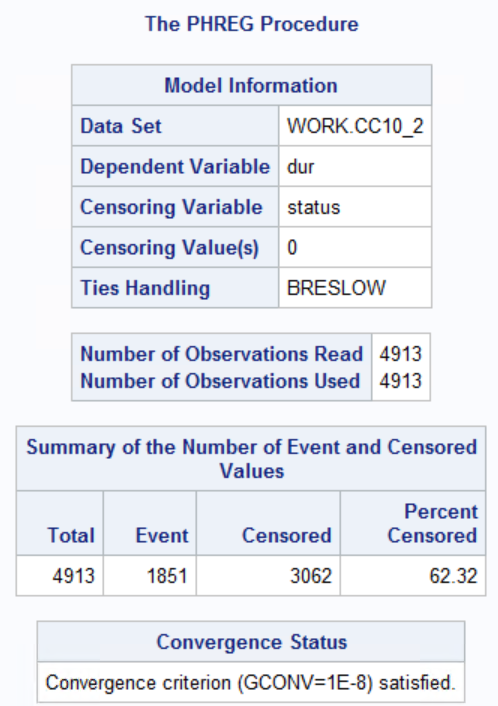
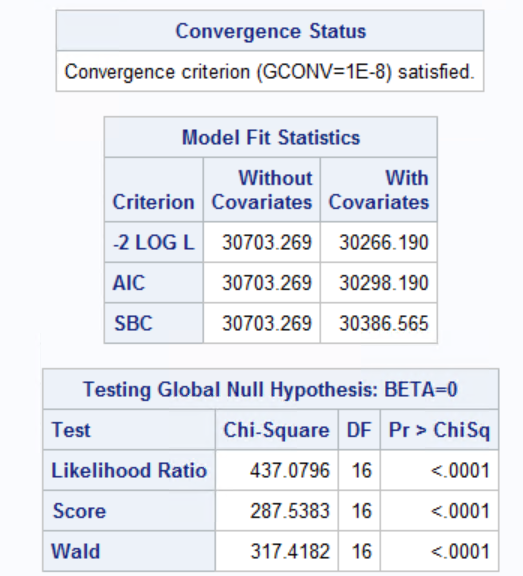
From the above analysis we can see that there are null values in the field as total record count is 7401 records and duration has only 4977 records.

Also looking at the above histogram showing the distribution of dur (duration) column, we can see the censoring in the data, there are far more cases with dur value of 37 than one would expect looking at the rest of the distribution.



*Censoring variable is Status, if status = 0 that means duration is censored otherwise it is not. There are no intercepts in the model as it is a characteristic of partial likelihood.*

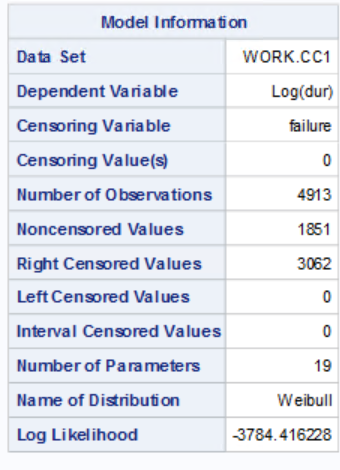
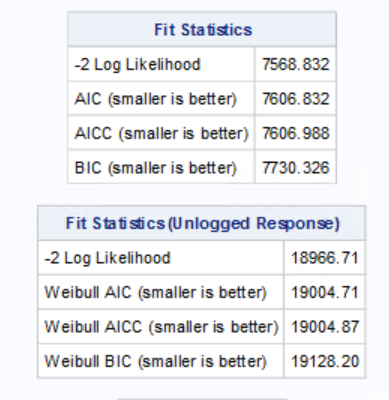
* Every $1000 increase in total transaction by the customer over 3-year period decreases the hazard of leaving the firm by 2.8%.
* Every $1000 increase in credit limit of the customer decreases the hazard of leaving the firm by 2.2%.
* The hazard of a platinum card customer leaving the firm is 47.8% more than the standard card customer.
* The hazard of a quantum card customer leaving the firm is 76.5 % more than the standard card customer.
* The hazard of customers with cards affiliated to professional organisations is 27% less than non-affinity card holders.
* The hazard of customers with cards affiliated to sports organisations is 22.2% less than non-affinity card holders.
* The hazard of customers with cards affiliated to Financial organisations is 28.8% less than non-affinity card holders.
* The hazard of customers with cards affiliated to University organisations is 28% less than non-affinity card holders.
* The hazard of customers with cards affiliated to commercial organisations is 39% less than non-affinity card holders.



Based on all the above three criteria, we can see that the values of the model with intercept and covariates is less than the intercept only criterion. Therefore, the model which we used is a better fit than the intercept only model.

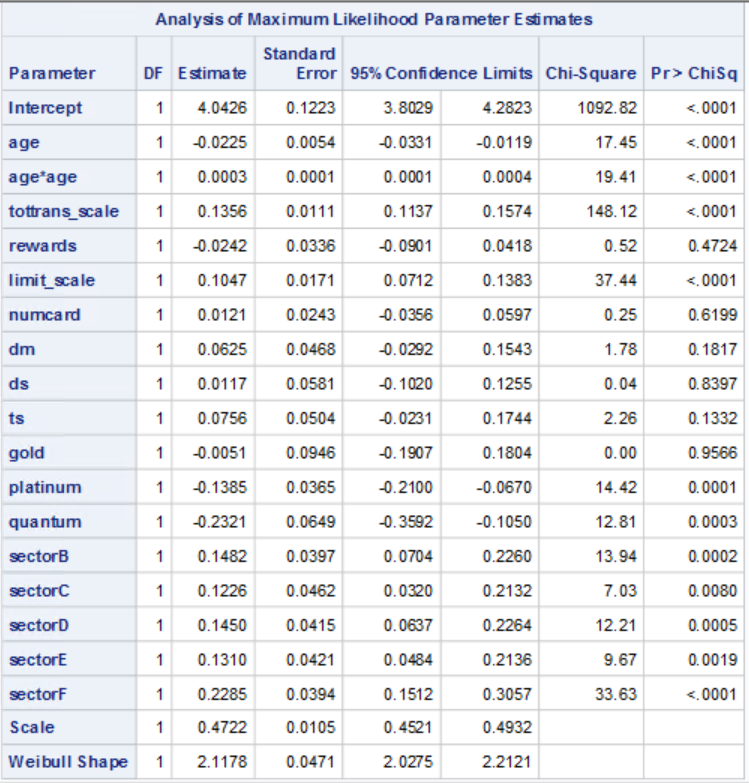
**4. Run a model using PROC LIFEREG with Weibull distribution.**

**Write a summary of the results. Focus on important effects, interpretation, model fit etc.**

From the above results we can see that the log likelihood for the Weibull distribution is –3784.42. The log-likelihood value can be used to compare the goodness of fit for nested models with different covariates, but with the same distribution.

The "Fit Statistics (Unlogged Response)" table is based on the maximum Weibull log likelihood using Duration as the response. The AIC, BIC, and AICC statistics in this table can be used to compare models with different covariates and different distributions, as long as the fit statistics for the models that you compare use Duration as the response.



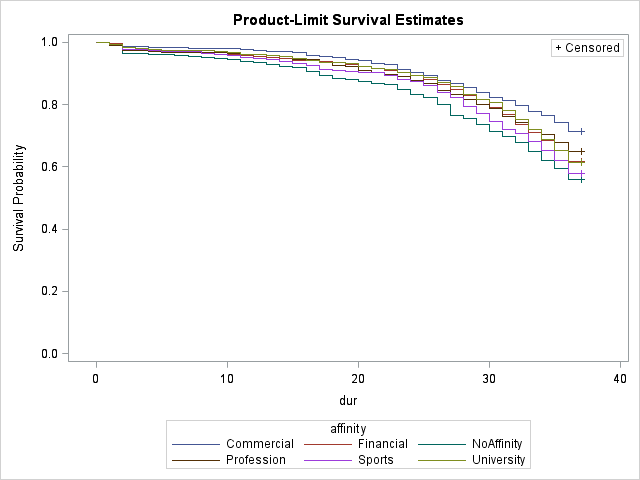
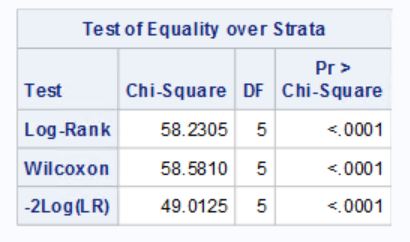
Failure variables captures the censoring effect. If failure = 0, that means the data is censored, otherwise not.

The important effects in the model are captured by:

1. tottrans\_scale: This variable is the total amount of transactions done by a customer. (scaled down to 1/10000). Higher transaction value corresponds to customers sticking to the firm for a longer period. For every $10,000 increase in the transaction value, the duration for which a customer stays with the firm increases by 13.56%
2. age has a non-linear relationship with the dependent variable. Increase in age causes the duration of customer subscription to go down.
3. limit\_scale: This is the credit limit of a customer (scaled down to 1/10000)/ Higher credit limit results in longer duration of a customer’s affiliation with the firm. For every $10,000 increase in the credit limit, the duration for which a customer remains in business with the firm increases by 10.47%
4. While there is not much difference in the duration of stay of standard and gold card holders, but platinum and quantum card holders leave the firm 13.85% and 23.21% earlier than standard card holders
5. Having affinity cards affects the duration of the customer’s stay with the firm positively. As compared to having no affinity card, there is an increase in stay duration of a customer by 14.82% for Professional organization affinity card, by 12.26 % for a Sports affinity card, 14.50% for a Financial institution affinity card, 13.10% for a University card and the highest, 22.85% for a commercial affinity card.

**5. Use PROC LIFETEST to test whether survivor function of affinity groups are significantly different from that of non-affinity groups.**

**What do you conclude?**



The commercial affinity group has the highest survival probability while having no affinity has the lowest survival probability.

Given is a table that consists of the approximate chi-square statistics, degrees of freedom, and *p*-values for the log-rank, Wilcoxon, and likelihood ratio tests. All three tests indicate strong evidence of a significant difference among the survival curves for the affinity groups (*p*<0.0001)

**Null Hypothesis**: The survival function of all groups is same.

Therefore, based on the above test, we can reject the null hypothesis to conclude that at least one of the groups is significantly different from one other group.